Spinal radiographic findings and nonspecific low back pain - A systematic review of observational studies

van Tulder, M.W.; Assendelft, W.J.J.; Koes, B.W.; Bouter, L.M.

published in
Spine
1997

DOI (link to publisher)
10.1097/00007632-199702150-00015

document version
Publisher's PDF, also known as Version of record

SPINE Volume 22, Number 4, pp 427–434

Spinal Radiographic Findings and Nonspecific Low Back Pain

A Systematic Review of Observational Studies

Maurits W. van Tulder, PhD, Willem J. J. Assendelft, MD, PhD, Bart W. Koes, PhD, and Lex M. Bouter, PhD

Study Design. A systematic review of published observational studies.

Objectives. To examine the causal relationship between radiographic findings and nonspecific low back pain.

Summary of Background Data. The causal relationship between radiographic findings and nonspecific low back pain still is controversial.

Methods. Two reviewers independently scored the methodologic quality of all relevant, available studies using a standardized set of criteria. The association between radiographic findings and nonspecific low back pain was expressed as an odds ratio with a corresponding 95% confidence interval.

Results. Degeneration, defined by the presence of disc space narrowing, osteophytes, and sclerosis, turned out to be associated with nonspecific low back pain with odds ratios ranging from 1.2 to 3.3. Spondylolysis and spondylolisthesis, spina bifida, transitional vertebrae, spondylosis, and Scheuermann's disease did not appear to be associated with low back pain. The validity scores of the observational studies ranged from 0% to 91% of the maximum score. Only two studies used a prospective design, and most studies lacked control for confounding, an appropriate test for nonspecific low back pain, and blinded assessment of radiographs and low back pain status.

Conclusions. There is no firm evidence for the presence or absence of a causal relationship between radiographic findings and nonspecific low back pain. [Key words: low back pain, observational studies, radiographs, systematic review] **Spine 1997;22:427–434**

Low back pain (LBP) is one of the major health problems in Western industrialized countries and a major cause of work absenteeism and disablement. The economic burden of LBP on society is alarming.[55] Although LBP is a rarely fatal, usually benign and self-limiting condition, it is a prominent problem in medical practice. Little is known about the etiology and pathogenesis of this complaint. The management of LBP remains symptomatic.

From the Institute for Research in Extramural Medicine, Faculty of Medicine, Vrije Universiteit, Amsterdam, The Netherlands.
Acknowledgment date: December 6, 1995.
First revision date: April 1, 1996.
Acceptance date: July 17, 1996.
Device status category: 1.

The management of LBP in primary care often includes a radiographic evaluation of the lumbar spine, which usually consists of anteroposterior and lateral views, supplemented with oblique and coned-down L5–S1 lateral views when indicated.[4,10,42] The main purpose of radiographic evaluation of the lumbar spine is to exclude the occurrence of LBP specifically caused by malignancies, infections, inflammatory spondyloarthropathies, and fractures. These may require specific therapies or substantially affect prognosis. Therefore, criteria have been proposed for the selective use of radiographs based on features of history-taking and physical examination ("red flags") suggesting specific LBP.[8] Despite the fact that specific LBP has a low prevalence in primary care, radiographs frequently are taken of patients with LBP.[5,53] This suggests that radiographs are requested when specific LBP is not suspected. In patients with nonspecific LBP, radiographs may be taken only for reassurance or at the patient's request.[28] Findings of degenerative, congenital, and postural abnormalities have been assumed to be associated with nonspecific LBP but the reliability of these findings is low.[6,7] The role of these radiographic abnormalities in the etiology of nonspecific LBP is unclear but may have consequences for preemployment screening or for the management of LBP.

To demonstrate a causal relationship between radiographic findings and nonspecific LBP from the available literature, it would be necessary to systematically review observational studies comparing individuals with and without LBP.[9] Obviously, radiographic findings must appear to be associated with nonspecific LBP. The most important evidence for causality of such an association is the strength of the study design or the extent of susceptibility to bias.[15] Therefore, a systematic review of observational studies should consider the internal validity of the studies by identifying potential sources of bias that may affect the outcome of these studies. The likelihood that an association is causal also is enhanced if other criteria are met, such as the temporality, strength, consistency, and biologic plausibility of the association.[15]

The use of systematic reviews of published data from observational studies has been a topic of debate.[21,43,48] We agree with Petitti[43] that the results of a systematic

review of observational studies cannot be considered definitive and should be interpreted with caution. We also agree with Greenland[21] that statistical pooling of heterogeneous observational studies by the use of a random-effects model does not explain the heterogeneity and is merely misleading. However, a systematic review in which the results are not pooled statistically, but in which the original studies and their results are systematically evaluated, would be a useful method of comparing and summarizing studies, identifying methodologic flaws, and determining potential biases.

In this study, we systematically reviewed the methodologic quality of all available, relevant observational studies to explore whether there is a causal relationship between abnormal findings on lumbar radiographs and nonspecific LBP.

■ Methods

Study Selection. A MEDLINE literature search was conducted for the period from January 1966 to September 1994 using the key words "diagnostic imaging," "spine," "backache," "back pain," "low back pain," "radiography," "roentgenograms," and "x-rays." In addition, the references cited in the selected papers were considered, regardless of the year of publication. A study was included in this review if 1) the study population included subjects with and without LBP, 2) at least one of the diagnostic tests was plain radiography, and 3) the study was published in English. Excluded were studies about specific LBP caused by malignancies, infections, inflammations, osteoporosis, or fractures and studies about flexion–extension radiographs. Nonhuman studies, cadaver studies, case reports (five subjects or less), abstracts, letters, and editorials also were excluded. One hundred ninety-four publications were selected from the MEDLINE search. An EMBASE search did not reveal any further publications. Finally, we sent our list of publications to six international experts in spinal radiography, asking them to check on the completeness of the list. Four responded that the list appeared complete, and no supplementary studies were suggested.

Methodologic Quality Assessment. Two reviewers (MWvT and WJJA) independently scored the methodologic quality of each study, according to a standardized set of predefined criteria (Table 1). The criteria list was based on previously published guidelines for assessing the methodologic quality of studies evaluating the accuracy of diagnostic tests[11,27,41] and on reviews concerning the evaluation of diagnostic tests.[22,23,30] The 14 criteria (A–N) referred to the study population, the assessment of radiographs, the assessment of LBP status, the blinded assessment of radiographs and LBP, and the analysis and data presentation. Initially, there was agreement between the two independent reviewers over the criteria in 361 (83%) of the 434 items scored. Disagreement (17%) mainly occurred because of reading and interpretation errors. In a consensus meeting, the disagreements between the two reviewers were discussed and resolved.

The criteria list contains items reflecting the internal and external validity of the studies. The criteria concerning the internal validity of the studies (items A, E, G–L, and N), which is the degree to which the results of a study are correct for the patients being studied,[15] were used to assess a hierarchical order of the methodologic quality (Table 2). Each of the criteria was assigned a maximum score of five points (for rating system, see the Appendix). Consequently, each study could score a maximum of 45 points. Studies scoring more than 50% of the maximum attainable score were, arbitrarily, considered to be of acceptable to good methodologic quality. Criteria B–D, F, and M, covering the external validity, which is the degree to which the results of a study are generalizable to other settings,[15] and the quality of reporting, were assessed for educational purposes (*i.e.*, the improvement of future research projects and publications).

Table 1. Methodologic Criteria for Observational Studies on the Association Between Spinal Radiographic Findings and Nonspecific Low Back Pain

Methodologic Criteria*	Score
Study population	
A Selection of study population	5
B Description of inclusion and exclusion criteria	5
C Description of potential confounders	5
Assessment of radiographs	
D Description of technique and equipment	5
E Definition of normal and abnormal result	5
F Reproducibility of test interpretation	5
Assessment of LBP status	
G Appropriate test for LBP	5
H Same test applied to all subjects	5
I Adequate follow-up period	5
Blinded assessment	
J Blinded assessment of radiographs	5
K Blinded assessment of LBP status	5
Analysis and data presentation	
L No missing values or description of missing values	5
M Presentation or reconstruction of 2 × 2 table	5
N Control for confounders	5

* For operationalization of the criteria; see Appendix.
LBP = low back pain.

Analysis. The odds ratios (ORs) and the 95% confidence intervals (CI) of the most prevalent diagnostic entities (*i.e.*, degenerative changes, spondylolysis and spondylolisthesis, spina bifida, transitional vertebrae, spondylosis, and Scheuermann's disease) were estimated using the Conference Interval Analysis-programme published by the British Medical Journal, London, 1989. We used the definitions of the diagnostic entities as given by the authors of the studies. Although the exact definition of degeneration varied among the studies, in general it was defined as the presence of disc space narrowing, osteophytes, or sclerosis. Biering-Sørensen et al[1] and Horal[24] distinguished degeneration from spondylosis, with the former defined by the presence of sclerosis or osteophytosis on the adjacent vertebrae or disc space narrowing and the latter by the presence of marginal osteophytes on the vertebral bodies. Hussar and Guller[26] defined spondylosis as the presence of osteophytes on the vertebral bodies, disc space narrowing, or narrowed intervertebral foramina.

■ Results

Methodologic Quality

We identified 35 publications meeting our inclusion criteria. Five studies were reported on twice, and because of

Table 2. Observational Studies on the Association Between Spinal Radiographs and Nonspecific Low Back Pain in Order of Validity Score

	Internal Validity										External Validity/Quality of Reporting				
Author	A 5	E 5	G 5	H 5	I 5	J 5	K 5	L 5	N 5	Score (%)*	B 5	C 5	D 5	F 5	M 5
Virta[56]	5	5	5	5	5	5	5	5	1	91	5	5	5	5	5
Symmons(c)[51,52]	5	5	5	5	5	5	5	5	0	89	5	4	5	5	5
Horal[24]	5	5	5	5	0	5	5	5	2	82	5	2	5	0	5
Frymoyer[17]	5	5	5	5	0	5	5	5	1	80	5	2	5	5	0
Biering-S.[1]	5	0	5	5	5	5	5	5	0	78	5	4	5	0	5
Fisk[14]	5	5	5	5	0	5	5	5	0	78	5	5	5	0	0
Riihimäki(c)[44,45]	5	5	5	5	0	0	5	5	5	78	5	2	5	5	5
Symmons(p)[51,52]	0	5	5	5	5	5	5	5	0	78	5	4	5	5	5
Wiikeri[57]	5	5	0	5	0	5	5	5	3	73	5	3	5	0	5
Lawrence[33]	5	5	5	5	0	5	5	0	2	71	5	2	0	0	5
Kellgren[29]	5	5	0	5	5	0	0	5	2	60	5	3	0	0	5
Magora[38]	5	5	0	5	0	5	0	5	2	60	0	2	0	0	5
Magora[39]	5	5	0	5	0	5	0	5	1	58	0	1	0	0	5
Riihimäki(p)[44,45]	5	5	5	5	5	0	0	0	1	58	5	5	5	5	5
Sairanen[47]	5	5	0	5	5	0	0	5	1	58	5	3	5	0	5
Hussar[26]	5	0	0	5	0	5	5	5	0	56	5	4	0	0	5
Magora[40]	5	5	0	5	0	5	0	5	0	56	0	0	0	0	5
Hult[25]	5	0	0	5	5	0	0	5	3	51	0	3	0	0	5
Magora[37]	2	5	0	5	0	0	5	0	4	47	0	4	0	0	5
Rowe[46]	5	0	5	5	0	0	0	0	2	38	0	2	5	0	5
LaRocca[31,32]	3	5	0	0	0	0	0	5	1	31	5	2	0	0	5
Torgerson[54]	0	5	0	0	0	0	0	5	2	27	0	2	5	0	5
Witt[58]	5	0	0	0	0	0	0	5	1	24	0	2	5	0	5
Bigos[2]	0	5	0	0	0	0	0	5	0	22	5	3	5	0	5
Fischer[13]	5	0	0	0	0	0	0	5	0	22	0	3	5	0	5
LeBoeuf[34]	0	5	0	0	0	0	0	5	0	22	0	2	0	0	5
Splithoff[50]	5	0	0	0	0	0	0	5	0	22	0	0	5	0	5
Brav[3]	3	0	0	0	0	0	0	5	0	18	0	1	5	0	5
Fullenlove[19]	0	0	0	0	0	0	0	5	0	11	0	0	0	0	5
Gibson[35,36]	0	0	0	0	0	0	0	5	0	11	0	3	0	0	5
Gillespie[20]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
% of Maximum	73	65	35	65	26	42	39	84	22		52	50	58	19	94

*Percentage of the maximum attainable validity score; if the validity scores are equal an alphabetical order is used.
(c) = cross-sectional part of the study. (p) = prospective part of the study.

completeness of information, the methodologic assessment was based on combining both publications.[13,16,31,32,35,36,44,45,51,52] Four publications resulting from the same large survey were scored separately. Each publication reported on a different radiographic finding and had a different study population.[37–40] The publications of Symmons et al[51,52] involved data from a cross-sectional and a prospective study, which were scored as separate studies. Consequently, 31 studies were included in our systematic review.

Table 2 presents all studies in a hierarchical order according to their methodologic quality. The methodologic quality of the studies varied strongly, with validity scores ranging from 0% to 91% of the maximum attainable score. Eighteen studies scored more than 50% and were considered to be of acceptable or good methodologic quality. Thirteen studies were of low methodologic quality. The most prevalent methodologic flaws (Table 2) concerned the criteria regarding control for confounding (item N; 22% of the maximum attainable score), the length of the follow-up period (item I; 26%), the appropriateness of the test for LBP (item G; 35%), the blinded assessment of LBP status (item K; 39%), and the blinded assessment of radiographs (item J; 42%). The methodologic criterion most frequently complied with was the description of missing values (item L; 84%).

Table 2 also shows that most of the criteria considered to be important for the external validity of the study or for the quality of the reporting were poorly met. The reproducibility of test interpretation (item F) scored only 19% of the maximum attainable score, whereas the description of inclusion–exclusion criteria (item B), potential confounders (item C), and technique and equipment (item D) scored 52%, 50%, and 58% of the maximum attainable score, respectively. Our criterion regarding the presentation of or the ability to reconstruct a 2 × 2 table (item M) was met in all but two studies.

Association Between Radiographic Findings and Nonspecific Low Back Pain

Because most studies did not report an OR, we had to estimate the ORs from the results presented in the orig-

Table 3. Odds Ratios (OR) and 95% Confidence Intervals (CI) for the Studies of the Association Between Radiographic Findings and Nonspecific Low Back Pain With a Validity Score of More Than 50%

		LBP		no LBP				
Author	%	Present	Absent	Present	Absent	OR	95% CI	Results*
Degeneration								
Symmons[51,52]†	89	130	106	92	149	1.99	1.38, 2.86	Positive
Symmons[51,52]‡	89	170	66	135	106	2.02	1.38, 2.96	Positive
Horal[24]	82	90	105	61	127	1.87	1.24, 2.83	Positive
Frymoyer[17]	80	45	151	19	77	1.21	0.66, 2.21	Positive
Biering-Sørensen[1]	78	115	243	71	237	1.58	1.12, 2.23	Positive
Riihimäki[44,45]	78	—	—	—	—	2.7§	1.8,4.0	Positive (sciatica)
Symmons[51,52]‖	78	40	66	43	106	1.49	0.88, 2.54	Not clear
Wiikeri[57]	73	39	28	42	100	3.32	1.83, 6.02	Positive
Lawrence[33]	71	462	320	360	390	1.56	1.28, 1.91	Positive
Kellgren[29]	60	55	18	77	45	1.79	0.94, 3.41	Positive
Sairanen[47]	58	139	35	51	41	3.30	1.89, 5.76	Not clear
Hult[25]	51	177	41	35	20	2.47	1.29, 4.71	Positive
Spondylosis								
Horal[24]	82	164	31	141	54	2.03	1.23, 3.33	Positive
Biering-Sørensen[1]	78	210	148	177	131	1.05	0.77, 1.43	No
Hussar[26]	56	26	69	89	316	1.34	0.80, 2.23	No
Spondylolysis/listhesis								
Virta[56]	91	34	32	12	14	1.24	0.50, 3.08	Positive (women)
Horal[24]	82	9	186	11	183	0.81	0.33, 1.99	No
Biering-Sørensen[1]	78	10	348	9	299	0.95	0.38, 2.38	No
Sairanen[47]	58	13	35	7	41	2.12	0.79, 5.64	Not clear
Magora[40]	56	132	516	164	212	0.33	0.25, 0.44	No
Hult[25]	51	6	226	3	129	1.14	0.28, 4.64	No
Spina bifida								
Horal[24]	82	15	180	21	173	0.69	0.34, 1.38	No
Magora[39]	58	284	526	232	212	0.50	0.40, 0.64	No
Transitional vertebrae								
Horal[24]	82	29	166	34	160	0.82	0.48, 1.41	No
Biering-Sørensen[1]	78	14	344	14	294	0.85	0.40, 1.82	No
Magora[38]	60	54	258	42	106	0.53	0.33, 0.84	No
m. Scheuermann								
Biering-Sørensen[1]	78	22	336	23	285	0.81	0.44, 1.49	No
Sairanen[47]	58	18	35	5	41	3.60	1.42, 9.14	Not clear

* Results according to the authors of the study (positive or no association).
† Cross-sectional data of 1975.
‡ Cross-sectional data of 1985.
§ Crude OR for sciatica; the OR adjusted for occupation, earlier back incidents, age, height, body mass index, and smoking was 1.9 (95% CI 1.2–2.9).
‖ Prospective data 1975–1985.

inal studies. If the study population included both sexes, we used the combined OR for men and women. We used the data of Virta et al[56] regarding LBP during the previous year and the combined data on prelysis, lysis, and listhesis from the data of Magora and Schwartz[40] to estimate the ORs of these studies. The scores for slight, moderate, and severe degeneration and for lumbago and sciatica were combined in estimating the OR from the data of Wiikeri et al.[57] We used only the prospective data of Symmons et al,[52] which involved the population without degeneration at the first examination.

The estimated ORs of the studies that were of acceptable or good methodologic quality are presented in Table 3. The ORs of the association between degeneration and LBP range from 1.21 to 3.32, with most 95% CIs not including 1, indicating a consistent and statistically significant positive association. Five of the seven studies reporting on an association between degeneration and nonspecific LBP with a methodologic score of 50% or less reported an OR around 1 with a 95% CI including 1 (data not shown). The ORs for spondylolysis and spondylolisthesis range from 0.33 to 2.12, with most 95% CIs including 1. Only one study reported a significantly higher frequency of spondylolysis and spondylolisthesis among men and women without nonspecific LBP.[40] Only a few studies of acceptable or good methodologic quality reported on the association between spina bifida, transitional vertebrae, spondylosis, and Scheuermann's disease. In general, these radiographic findings do not appear to be associated with nonspecific LBP.

■ Discussion

In this study, we systematically reviewed all available observational studies. Comparing radiographic findings of subjects with and without nonspecific LBP in a "best-evidence synthesis" approach, we based our conclusions on the studies with the highest methodologic quality.[49] The methodologic quality of the studies included in our systematic review varied widely, from 0% to 91% of the maximum attainable validity score. We considered stud-

ies scoring more than 50% to be of acceptable or good methodologic quality and studies scoring 50% or less to be of low methodologic quality. The results of the 18 higher-quality studies indicated that degeneration is associated with nonspecific LBP. Other radiographic findings are not. Even in the best quality studies, we encountered some common methodologic flaws concerning the control for confounding, the length of the follow-up period, the appropriateness of the test for LBP, the blinded assessment of LBP status, and the blinded assessment of radiographs, which all may have resulted in biased outcomes.

A potential limitation of our systematic review was the literature search, which may have introduced a publication bias for several reasons. First, some relevant studies might have been missed because they used other key words or had unclear abstracts. We tried to solve this problem by using a broad search strategy (see Methods section for the key words used) and by checking the original article if the abstract raised any doubt. Second, not all published studies are indexed in databases, and we could have missed some relevant studies that were published in nonindexed journals. We tried to avoid this problem by retrieving studies from two international databases (MEDLINE and EMBASE), by screening all reference lists of reviews and original publications, and by sending our selection of studies to experts in the field. Third, we did not make an effort to identify unpublished studies. In general, unpublished studies are more likely to have results that are not statistically significant. Fourth, we selected studies published before 1966 by screening the reference lists of relevant publications, and studies reporting a positive or statistically significant association probably are more likely to be referred to in other publications. The identification of all relevant studies is crucial to the validity of systematic reviews; therefore, adequate indexing of published studies and registration of unpublished studies should be aimed at to reduce the possibility of publication bias.[12]

Only a few studies used an appropriate test for LBP and described the diagnostic criteria for interpretation of the test. We defined history taking and physical examination or the use of a standardized questionnaire as appropriate tests if the presence, duration, and severity of pain were assessed (see the Appendix). The use of an inappropriate test may lead to substantial misclassification. If the test is applied uniformly to individuals with and without LBP, which was the case in all 18 studies with a methodologic score of more than 50%, but in only two of the 13 studies with a score equal to or less than 50%, the misclassification will be similar for both groups (nondifferential), resulting in an underestimation of the true ORs.

The assessment of radiographs and the assessment of the LBP status was not blinded in 58% and 61% of the studies, respectively. If blinding does not take place, knowledge of the LBP status unavoidably influences the observer in the assessment of the radiographs, which may occur in case–control studies. The results of the radiographic examination also might influence the patient in reporting about the LBP, which may occur in cohort studies. Blinding of the observers and the study population is not difficult when examining the association between radiographic findings and nonspecific LBP, and future studies should incorporate blinding to prevent this type of bias, which results in an overestimation of the true ORs.

Selection bias did not appear to be a major problem in most studies. The selection of the study population (item A) scored 73% of the maximum attainable score and was described and considered as adequate in all but one of the better quality studies.

Another potential source of bias, confounding, was not examined in most of the studies. Confounding occurs when two factors or processes are associated or "travel together," and when the effect of one is confused with or distorted by the effect of the other.[15] To be considered a confounder in the estimated association between radiographic findings and nonspecific LBP, a variable must be a risk factor for nonspecific LBP, must be associated with the radiographic findings at issue, and must not be an intermediate factor in the causal pathway between radiographic findings and nonspecific LBP. Because confounding is an important issue in observational studies and because it may lead to an underestimation or overestimation of the true ORs, it should be controlled for in the analysis. Prospective studies in which confounding is evaluated and controlled for are necessary to evaluate the association between radiographic findings and nonspecific LBP.

Only two studies, those of Riihimäki et al[44,45] and Symmons et al,[51,52] used this type of prospective design. Most of the reviewed studies were case–control studies in which the LBP status and the radiographic findings were assessed at the same point in time or in which the radiographs were taken at the time of investigation and related to nonspecific LBP in the past. One of the criteria for causality of an association is temporality, which means that causes should precede effects. If it is hypothesized that the radiographic findings are related to the cause of nonspecific LBP, the radiographic examination should precede the occurrence of nonspecific LBP in time and definitely should not be studied in relation to nonspecific LBP in the past. The latter can only examine the hypothesis that nonspecific LBP produces radiographic abnormalities, which no one will endorse.

The results of our systematic review show that spondylolysis and spondylolisthesis, spina bifida, transitional vertebrae, spondylosis, and Scheuermann's disease do not seem to be associated with nonspecific LBP. However, degeneration does seem to be associated with nonspecific LBP, with ORs ranging from 1.21 to 3.32. As we have outlined earlier, several potential sources of bias could be identified that may have led to an underestima-

tion or overestimation of the true association. Even if there is a true association between radiographic findings and degeneration, the strength of the association as expressed by the ORs is not convincing. Further, the temporality of the association was overlooked in most studies. Therefore, we conclude that no firm evidence exists for the presence or absence of a causal relationship between radiographic findings and nonspecific LBP.

References

1. Biering-Sørensen F, Hansen FR, Schroll M, Runeborg O. The relation of spinal X-ray to low-back pain and physical activity among 60-year old men and women. Spine 1985;10: 445–51.
2. Bigos SJ, Hansson T, Castillo RN, Beecher PJ, Wortley MD. The value of preemployment roentgenographs for predicting acute back injury claims and chronic back pain disability. Clin Orthop 1992;283:124–9.
3. Brav EA, Bruck S, Fruchter JM. A roentgenologic study of low back and sciatic pain. AJR Am J Roentgenol 1942;48:39–46.
4. Butt WP. Radiology for back pain. Clin Radiol 1989;40: 6–10.
5. Carey TS, Evans A, Hadler N, Kalsbeek W, McLaughlin C, Fryer J. Care-seeking among individuals with chronic low back pain. Spine 1995;20:312–7.
6. Coste J, Paolaggi JB, Spira A. Reliability of interpretation of plain lumbar spine radiographs in benign, mechanical low-back pain. Spine 1991;16:426–8.
7. Deyo RA, McNiesh LM, Cone RO. Observer variability in the interpretation of lumbar spine radiographs. Arthritis Rheum 1985;28:1066–70.
8. Deyo RA, Diehl AK. Lumbar spine films in primary care: Current use and effects of selective ordering criteria. J Gen Intern Med 1986;1:20–5.
9. Deyo RA. Early diagnostic evaluation of low back pain. J Gen Intern Med 1986;1:328–38.
10. Deyo RA. Diagnostic imaging procedures for the lumbar spine. Ann Intern Med 1989;111:865–7.
11. Deyo RA, Haselkorn J, Hoffman R, Kent DL. Designing studies of diagnostic tests for low back pain or radiculopathy. Spine 1994;19(Suppl):S2057–65.
12. Dickersin K. The existence of publication bias and risk factors for its occurrence. JAMA 1990;263:1385–9.
13. Fischer FJ, Friedman MM, VanDemark RE. Roentgenographic abnormalities in soldiers with low back pain: A comparative study. AJR Am J Roentgenol 1958;79:673–6.
14. Fisk JW, Baigent ML, Hill PD. Scheuermann's disease: Clinical and radiological survey of 17 and 18 year olds. Am J Phys Med 1984;63:18–30.
15. Fletcher RH, Fletcher SW, Wagner EH. Clinical Epidemiology: The Essentials. 2nd ed. Baltimore: Williams & Wilkins, 1988.
16. Friedman MM, Fischer FJ, VanDemark RE. Lumbosacral roentgenograms of one hundred soldiers: A control study. AJR Am J Roentgenol 1946;55:292–8.
17. Frymoyer JW, Newberg A, Pope MH, Wilder DG, Clements J, MacPherson BJ. Spine radiographs in patients with low-back pain. J Bone Joint Surg [Am] 1984;66:1048–55.
18. Frymoyer JW, Phillips RB, Newberg AH, MacPherson BV. A comparative analysis of the interpretations of lumbar spinal radiographs by chiropractors and medical doctors. Spine 1986; 11:1020–3.
19. Fullenlove TM, Williams AJ. Comparative roentgen findings in symptomatic and asymptomatic backs. Radiology 1957;68:572–4.
20. Gillespie HW. The significance of congenital lumbo-sacral abnormalities. Br J Radiol 1949;22:270–5.
21. Greenland S. Can meta-analysis be salvaged? Am J Epidemiol 1994;140:783–7.
22. Hoffman RM, Kent DL, Deyo RA. Diagnostic accuracy and clinical utility of thermography for lumbar radiculopathy: A meta-analysis. Spine 1991;16:623–7.
23. Hoogen HMM van den, Koes BW, Eijk JThM van, Bouter LM. On the accuracy of history, physical examination, and erythrocyte sedimentation rate in diagnosing low back pain in general practice: A criteria-based review of the literature. Spine 1995;20:318–27.
24. Horal J. The clinical appearance of low back disorders in the city of Gothenburg, Sweden: Comparison of incapacitated probands with matched controls. Acta Orthop Scand Suppl 1969;118:1–73.
25. Hult L. The Munkfors investigation. Acta Othop Scand Suppl 1954;16:1–75.
26. Hussar AE, Guller EJ. Correlation of pain and the roentgenographic findings of spondylosis. Am J Med Sci 1956;232: 518–27.
27. Irwig L, Tosteson ANA, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 1994;120:667–76.
28. Kaplan DM, Knapp M, Romm FJ, Velez R. Low back pain and X-ray films of the lumbar spine: A prospective study in primary care. South Med J 1986;79:811–4.
29. Kellgren JH, Lawrence JS. Rheumatism in miners: Part II. X-ray study. Br J Ind Med 1952;9:197–207.
30. Kent DL, Haynor DR, Larson EB, Deyo RA. Diagnosis of lumbar spinal stenosis in adults: A metaanalysis of the accuracy of CT, MR, and myelography. AJR Am J Roentgenol 1992; 158:1135–44.
31. LaRocca H, MacNab I. Value of pre-employment radiographic assessment of the lumbar spine. Can Med Assoc J 1969;101:383–8.
32. LaRocca H, MacNab I. Value of pre-employment radiographic assessment of the lumbar spine. Ind Med 1970;39: 253–8.
33. Lawrence JS. Disc degeneration: Its frequency and relationship to symptoms. Ann Rheum Dis 1969;28:121–37.
34. Leboeuf C, Kimber D, White K. Prevalence of spondylolisthesis, transitional anomalies and low intercrestal line in a chiropractic patient population. J Manipulative Physiol Ther 1989;12:200–4.
35. Libson E, Bloom RA, Dinari G. Symptomatic and asymptomatic spondylolysis and spondylolisthesis in young adults. Int Orthop 1982;6:259–61.
36. Libson E, Bloom RA, Dinari G, Robin GC. Oblique lumbar spine radiographs: Importance in young patients. Radiology 1984;151:89–90.
37. Magora A, Schwartz A. Relation between the low back pain syndrome and X-ray findings: 1. Degenerative osteoarthritis. Scand J Rehabil Med 1976;8:115–25.
38. Magora A, Schwartz A. Relation between the low back pain syndrome and X-ray findings: 2. Transitional vertebra (mainly sacralization). Scand J Rehabil Med 1978;10:135–45.

39. Magora A, Schwartz A. Relation between the low back pain syndrome and X-ray findings: 3. Spina bifida occulta. Scand J Rehabil Med 1980a;12:9–15.
40. Magora A, Schwartz A. Relation between the low back pain syndrome and X-ray findings: 4. Lysis and olisthesis. Scand J Rehabil Med 1980b;12:47–52.
41. Mulrow CD, Linn WD, Gaul MK, Pugh JA. Assessing quality of a diagnostic test evaluation. J Gen Intern Med 1989; 4:288–95.
42. Pelz DM, Haddad RG. Radiologic investigation of low back pain. Can Med Assoc J 1989;140:289–95.
43. Petitti DB. Of babies and bathwater. Am J Epidemiol 1994;140:779–82.
44. Riihimäki H, Wickström G, Hänninen K, Mattsson T, Waris P, Zitting A. Radiographically detectable lumbar degenerative changes as risk indicators of back pain. Scand J Work Environ Health 1989a;15:280–5.
45. Riihimäki H, Wickström G, Hänninen K, Luopajärvi T. Predictors of sciatic pain among concrete reinforcement workers and house painters: A five-year follow-up. Scand J Work Environ Health 1989b;15:415–23.
46. Rowe ML. Low back pain in industry. J Occup Med 1969; 11:161–9.
47. Sairanen E, Brüshaber L, Kaskinen M. Felling work, low-back pain and osteoarthritis. Scand J Work Environ Health 1981;7:18–30.
48. Shapiro S. Meta-analysis/shmeta-analysis. Am J Epidemiol 1994;140:771–8.
49. Spitzer WO. Meta-meta-analysis: Unanswered questions about aggregating data. J Clin Epidemiol 1991;44:103–7.
50. Splithoff CA. Lumbosacral junction. Roentgenographic comparison of patients with and without backaches. JAMA 1953;152:1610–3.
51. Symmons DPM, van Hemert AM, Vandenbroucke JP, Valkenburg HA. A longitudinal study of back pain and radiological changes in the lumbar spines of middle aged women: I. Clinical findings. Ann Rheum Dis 1991a;50:158–61.
52. Symmons DPM, van Hemert AM, Vandenbroucke JP, Valkenburg HA. A longitudinal study of back pain and radiological changes in the lumbar spines of middle aged women: II. Radiographic findings. Ann Rheum Dis 1991b;50:162–6.
53. Szpalski M, Nordin M, Skovron ML, Melot C, Cukier D. Health care utilization for low back pain in Belgium: Influence of sociocultural factors and health beliefs. Spine 1995;20:431–42.
54. Torgerson WR, Dotter WE. Comparative roentgenographic study of the asymptomatic and symptomatic lumbar spine. J Bone Joint Surg [Am] 1976;58:850–3.
55. Tulder MW van, Koes BW, Bouter LM. A cost-of-illness study of back pain in the Netherlands. Pain 1995;62:232–40.
56. Virta L, Rönnemaa T, Osterman K, Aalto T, Laakso M. Prevalence of isthmic lumbar spondylolisthesis in middle-aged subjects from eastern and western Finland. J Clin Epidemiol 1992;45:917–22.
57. Wiikeri M, Nummi J, Riihimäki H, Wickström G. Radiologically detectable lumbar disc degeneration in concrete reinforcement workers. Scand J Work Environ Health 1978; 1(Suppl):47–53.
58. Witt I, Vestergaard A, Rosenklint A. A comparative analysis of X-ray findings of the lumbar spine in patients with and without lumbar pain. Spine 1984;9:298–300.

Address reprint requests to

Maurits W. van Tulder, PhD
Institute for Research in Extramural Medicine
Faculty of Medicine
Vrije Universiteit
Van der Boechorststraat 7
1081 BT Amsterdam
The Netherlands

■ Appendix. Operationalization of the Methodologic Criteria

Each criterion must be applied independently of the other criteria.

A. Positive if sampling or selection of the study population was described and considered adequate. In cohort studies: sampling before low back pain (LBP) has occurred (5 points). In case–control studies: selection of cases (2 points) and controls (3 points) before radiographs are taken.
B. Positive if the inclusion and exclusion criteria were adequately described to enable replication of the study.
C. Positive if potential confounders were described (frequency or mean ± standard deviation (SD) given for populations with and without LBP; 1 point each):
   1. Age
   2. Gender
   3. Work status (heavy physical work, static physical load, monotonous work)
   4. Smoking
   5. Obesity (weight or body mass index)
D. Positive if the technique was described adequately to enable replication of the study (anteroposterior, lateral or oblique view; standing or lying subjects; anatomic structures involved).
E. Positive if diagnostic criteria were defined adequately (*e.g.*, degenerative index).
F. Positive if the interobserver or intraobserver variability (depending on the study at issue) in the interpretation of the radiographs was described.
G. Positive if the LBP status was appropriately determined in a standardized manner. History taking, physical examination, or the use of a questionnaire to assess the presence or absence, the duration, severity, and the moment of occurrence (past/present/future) of pain and symptoms should be described clearly.
H. Positive if determination of the LBP status was uniformly applied to all subjects.
I. Period between radiographic examination and assessment of LBP had to be long enough for the LBP to occur; no follow-up period necessary for congenital

disorders, but 5 years for all noncongenital (degenerative, acquired, postural) disorders.

J. Positive if the interpretation of the radiographs was applied strictly without knowledge of the LBP status.

K. Positive if the determination of the LBP status was applied strictly independent of the results of the radiographs.

L. Positive if there were no missing values or if missing values weredescribed adequately, including loss to follow-up and drop-outs.

M. Positive if odds ratios (ORs) were given or if data were presented in a 2 × 2 table (or such a table could be constructed, to enable calculation of the ORs).

N. Positive if confounding was controlled for in the analysis (1 point for each confounder of item C).

■ Point of View

Jeremy Fairbank, MD
Nuffield Orthopaedic Centre
Oxford, UK

The group has published a series of important systematic reviews, of which this is the latest. They have been unable to demonstrate any relationship between radiographic findings and a diagnosis of non-specific low back pain in the papers that they studied. In our center, as magnetic resonance imaging (MRI) has become more available and the cost of a basic examination of the lumbar spine has risen, radiologists have discontinued offering plain radiographs and have made MRI available to general practitioners for the examination of their patients with back pain. Clinicians who read *Spine* are well aware of the rarity of "normal" MRI scans in this group of patients. The radiologist's report often makes alarming reading to general practitioners and their patients when both parties are seeking reassurance that nothing serious is amiss. At the same time, MRI probably is much more effective than plain radiographs in detecting the causes of "specific" back pain in this population. This paper confirms that in developed medical care systems there is little place for the plain radiographic examination in the investigation of non-specific low back pain. However, in many countries this is the only imaging system that is readily available. This study is important in these places. In all circumstances, skilled clinical methods should be promoted and investigation only requested when based on sound guidelines (see authors' reference 8). Our patients' enthusiasm for radiologic imaging needs to be tempered by public education, and general practitioners' skills need to be enhanced by specific training in low back pain and its investigation.